USAID
FROM THE AMERICAN PEOPLE

PERFORMANCE & IMPACT EVALUATION (P&IE) OF THE USAID/UGANDA SCHOOL HEALTH AND READING PROGRAM: RESULT 1 INTERVENTIONS

Impact Evaluation Report for Cluster 1, Year 2

30 APRIL 2015

This publication was produced at the request of the United States Agency for International Development. It was prepared independently by NORC at the University of Chicago.

PERFORMANCE & IMPACT EVALUATION (P&IE) OF THE USAID/UGANDA SCHOOL HEALTH AND READING PROGRAM: RESULT 1 INTERVENTIONS

IMPACT EVALUATION REPORT FOR CLUSTER 1, YEAR 2

30 April 2015

PN 7384; USAID Contract N0: AID-617-C-12-00006

PRESENTED TO:

USAID/Uganda
Joseph Mwangi

PRESENTED BY:

NORC at the University of Chicago
4350 East-West Highway, 8th Floor
Bethesda, MD 20814
Telephone: (301) 634-9413
Fax: (301) 634-9301

DISCLAIMER

CONTENTS

INTRODUCTION

NORC at the University of Chicago, in collaboration with Panagora, is conducting the Impact and Performance Evaluation of USAID/Uganda's School Health and Reading Program (SHRP), implemented by RTI. RTI (referred to as the implementing Partner or IP in the rest of the document) is implementing SHRP as two separate activities: (1) Activities related to Result 1, Improved Early Grade Reading and Transition to English, and (2) activities related to Result 2, Improved HIV/AIDS Knowledge, Attitudes and Practice. This report focuses on activities related to Result 1 (early grade reading) and presents the findings of the impact evaluation of the project on Cluster 1 students for two years of intervention.[1]

The report begins with a description of SHRP's Improved Early Grade Reading and Transition to English (Result 1) activities, followed by a description of the evaluation design, an analysis of baseline balance, and findings from the impact analysis.

A. PROJECT DESCRIPTION

For its Improved Early Grade Reading and Transition to English (Result 1) intervention, SHRP focuses on the nexus of language, pedagogy, and instructional materials to significantly improve students' early grade reading and literacy scores, as well as bring to scale a "Ugandan led 'reading policy'" (RTI International, 2012, p. 1). To this end, the program will build institutional capacity, support policy development and help institutionalize the training, support structures and policies necessary for sustainability. The intervention works at multiple levels. In early grade reading (Result 1), they consist of the following support to the MoES:

- At the school level, the intervention is comprised of training teachers in early grade literacy instruction using students' mother-tongue in P1-P3 and with a transition to English in P4; distribution of textbooks and instructional materials and teacher guides in local languages and English; support supervision provided to teachers; trainings in leadership for head teachers; and reading competitions and community awareness activities (reading awareness days, literacy week).
- At the district level, training for district education officials is raising awareness and building district-level support for the new curriculum.
- At the national level, MoES systems and pedagogical and language frameworks aim to strengthen and support mother-tongue based EGR and transition to English. The intent is to support the strengthening of policies related to reading; increase advocacy for reading at multiple levels (e.g. student, teacher, school, district, and national); and generate and use data for programmatic decision making.

Together, these interventions are expected to improve the instruction and learning environment of students and eventually lead to improved literacy skills.

Teacher trainings under Result 1 are implemented through Coordinating Center Tutors (CCTs), who are school support workers in charge of monitoring education quality within their Coordinating Centers (CC). Under the district education structure, each CCT is responsible for a certain number of schools within a district (one district typically has multiple CCTs). The CCTs selected for the intervention receive training directly from SHRP and, in turn, deliver teacher training and program support in their schools, thus following a Training of Trainers (TOT) model.

Teacher trainings at the school level focus on pedagogy with an emphasis on using structured lesson plans and learner books. These lesson plans provide teachers with a practical step-by-step process for implementing the transitional bilingual approach mandated by the Ugandan EGR policy. In addition, SHRP is developing materials to support early grade reading. These materials are being adapted in order to take into account the different needs of learners at different stages of cognitive and academic development, and the linguistic characteristics of the different local languages, rather than be translated directly from one language to another. Furthermore, in order to develop these materials, SHRP works with MoES and Local Language Boards (LLBs) to standardize orthographies of the target languages. All materials follow the same general pedagogical framework to facilitate guidelines for textbook development and teacher training. Teachers receive these instructional materials in their local language of instruction and English. Teachers also receive ongoing monitoring support supervision from SHRP staff who observe classes and provide constructive feedback.

Additionally, at the district level, workshops and training for district education officials are serving as a forum for raising awareness and building district-level support for the new curriculum. Communication campaigns directed at Members of Parliament and district leaders are also organized.

Finally, at the national level, SHRP is working with MoES and the Sector Policy Management Working Group to develop a Uganda-specific reading strategy, which will include policies in the areas of Local Language Board development, textbook development, printing as well as Special Needs. In particular, SHRP is working to develop a national literacy strategy, national reading standards and benchmarks, as well as to harmonize reading assessment efforts with the Uganda National Examination Board (UNEB). SHRP is also working to strengthen Local Language Boards (LLB) and to raise awareness about special needs education. SHRP is also assisting the MoES in advocating for reading outside the classroom. Together with MoES, SHRP aims to raise awareness of local language development, reading instructions and special needs learners by using national communication campaigns through mass media and mobilizing local communities.

Initially, the Implementing Partner (IP) had planned to develop variations of the Result 1 Cluster 1 intervention and use experimental approaches to identify the most effective intervention to scale up. In addition to distribution of textbooks and parental and community awareness, three slightly different interventions, or "treatment arms" planned for the first year of the project included: (1) Basic Program: teacher trainings alone; (2) Basic Program + manpower support: teacher trainings with a certain number of CCT visits to schools per year; and (3) Basic Program + SMS support: teacher trainings with SMS support by CCTs. However, during the implementation phase, the IP decided that it was not feasible to implement these different treatment arms during the first year and, hence, only started implementation of these different treatment arms in the second year of the project.

In total, SHRP will work in 12 local languages, developing instructional materials for each language and providing training and in-service support to teachers in areas where these 12 languages are spoken and used for mother-tongue based instruction in the early grades. The intervention is implemented following a pipeline roll-out, with the intervention targeting the districts associated with the first four local languages starting in Year 1 (Cluster 1), then targeting the districts associated with an additional four local languages starting Year 2 (Cluster 2), and finally targeting the districts associated with the final four local languages starting in Year 3 (Cluster 3). As such, Cluster 1 students will receive the intervention for a total of 4 years, Cluster 2 students for 3 years and Cluster 3 students for 2 years.

[1] At the time of submission of this report, the Result 1 team had collected data for: Round 1 for Cluster 1 (Feb-Mar 2013), Round 2 for Cluster 1 (Oct 2013), Round 3 for Cluster 1 (Oct 2014), Round 1 for Cluster 2 (Feb 2014) and Round 2 for Cluster 2 (Oct 2014), and the Result 2 team had collected baseline KAP data, but no follow-up data. Therefore, this report focuses on the impact of Result 1 on Cluster 1 students, using Cluster 1 Round 1 and Round 3 data. The results of the impact evaluation for Cluster 2 students is presented in a separate report. The first impact evaluation for Result 2 will only be carried out after the first follow-up data collection scheduled for mid-2015.

Over the 5 years of the project, SHRP will target the following local languages and associated districts:

Cluster	Local Language	Region	Districts
1	Luganda	Central	Wakiso, Gomba
	Runyankore/Rukiga	South West	Kiruhura, Bushenyi, Kabale
	Ateso	Eastern	Kumi, Katakwi, Serere
	Leblango	Northern	Apac, Lira, Kole
2	Runyoro/Rutoro	Mid-Western	Masindi, Kyenjojo, Kbarole
	Acholi	Mid-Northern	Gulu, Pader, Kitgum
	Lugbarati	West Nile	Arua
	Lumasaaba	Mid-Eastern	Mbale, Sironko, Manafwa
3	Lugwere	Mid-Eastern	Budaka, Pallisa, Kibuku
	Nkarimojong	North East	Nakapiripirit, Napak, Abim
	Lukhonzo	Mid-Eastern	Kasese
	Lusoga	East Central	Iganga and Kamuli

A.1 SHRP IN YEAR 1 AND YEAR 2 FOR CLUSTER 1

This report focuses on the impact of SHRP after two years of intervention for Cluster 1. For Cluster 1, SHRP worked in 4 local languages (Luganda, Runyankore/Rukiga, Ateso, and Leblano) in 11 districts[2]. By the end of Year 1, two teacher trainings had been conducted (May and September 2013). However, there were several delays in the development and distribution of instructional materials and textbooks. These materials and textbooks were not distributed until August/September 2013, i.e. approximately 6 months after the start of the school year. In Year 2, two teacher trainings on early grade reading (January and May 2014) and a teacher training on leadership (August 2013) were conducted. Materials for P2 where distributed on time and were available at the beginning of the academic year. In Year 2, SHRP distributed 151,508 Cluster 1 pupil primers and 3837 teacher's guides. SHRP also worked with International Book Bank and Books for Africa to select supplementary readers from various publishers and distributed them to schools in Wakiso, Gomba, and Kiruhura districts. Teachers from those schools received training in library management in June-July 2014.

In Year 2, SHRP also developed local language pupil primers and accompanying teacher guides for Primary 3 Cluster 1 languages as well as in English. The English materials focus on speaking, listening, reading and writing in English as a Second Language to facilitate transition to English as medium of instruction in Primary 4. The program also supported the Special Needs Education unit of the MoES to develop programs and activities, such as development of instructional materials that raise teachers' awareness of special needs students.

In Year 2, SHRP organized one communication campaign targeted at Members of Parliament and district leaders. Field Assistants also organized advocacy meetings in 60 communities. Program staff, education officials and field assistants initiated community mobilization and advocated for literacy instruction at Parent Teacher Association and School Management Committee meetings.

B. EVALUATION DESIGN

To assess the impact of the School Health and Reading Program on Cluster 1 students, NORC is using a combination of an experimental (randomized controlled trial, or RCT) and quasi-experimental (matched comparisons) design. This mixed-method design allows us to estimate the combined effects of the district- and school-level interventions that comprise the School Health and Reading Program; it also allows us to isolate the effects of the school-level intervention.

An impact evaluation (IE) is conducted to assess the causal effect of a specific intervention on a set of outcomes. It allows us to attribute changes in an outcome to a specific intervention or set of interventions by answering the counterfactual question "What would have happened to program participants in the absence of the intervention?" Ideally, this is done by observing the same program participants both with and without the intervention at the same point in time. Of course, this is not possible; at any given time, a participant either receives the intervention or not. Therefore, we can never directly observe the counterfactual and, instead, need to create a comparison group to serve as the counterfactual. Identifying a credible comparison group is a critical aspect of an impact evaluation.

The ideal comparison group stems from the use of experimental methods in which eligible participants are randomly assigned to receive the intervention or not. Randomization ensures that, on average, characteristics of the treatment and control groups are statistically identical, with the only difference being their participation in the intervention. In this case, any measured difference in outcomes between the groups over time can be attributed to the program. When random assignment is not possible, quasi-experimental methods, such as statistical matching, are used to establish a comparison group.

The impact evaluation design for the SHRP Result 1 Cluster 1 intervention uses both the random assignment of schools to treatment and control groups within SHRP intervention districts (experimental design) and the selection of matched comparison districts (quasi-experimental) in which SHRP is not operating. The experimental design allows us to isolate the effect of school-level interventions from district-level interventions, while the inclusion of non-intervention districts allows us to measure the impact of the district level interventions and the combined district+school level intervention package.

B.1 DISTRICT-LEVEL COMPARISON GROUP

Starting in the first year of the program, the IP implemented the literacy intervention with Cluster 1 students in 11 Districts located in 4 different language areas (Table 1). These 11 districts were chosen by the IP and MoES, and were not part of the evaluation design. Therefore randomization at the district level was not possible. However, we were able to select comparison districts that are similar in some key characteristics to the treatment districts. Although we had intended to pair a comparison district to each treatment district, budget and logistical restrictions expressed by the IP resulted in the selection of only one district per language adding to a total of four comparison districts to compare against 11 treatment districts.

The four control districts were selected by matching non-intervention and intervention districts in a specific language area according to district characteristics such as NAPE 2011 results on P3 proficiency in oral reading, P3 proficiency in literacy in English, and P6 proficiency in literacy.[3] Because we were matching only one comparison district to more than one intervention district, we computed a weighted average of treatment districts' proficiency scores, where the weights are proportional to the number of schools participating in the program during the first year. Through this matching process, we selected four control districts - Buikwe, Ngora, Otuke, and Ibanda (Table 1).

[2] Throughout the life of the project, SHRP will be working in a total of 12 local languages (4 languages in the first year, 4 additional languages in the second year, and 4 additional languages in the third year).

[3] Unfortunately, no information about HIV and AIDS knowledge, attitude and practices was available at the time of matching districts.

Table 1. Treatment and Comparison Districts

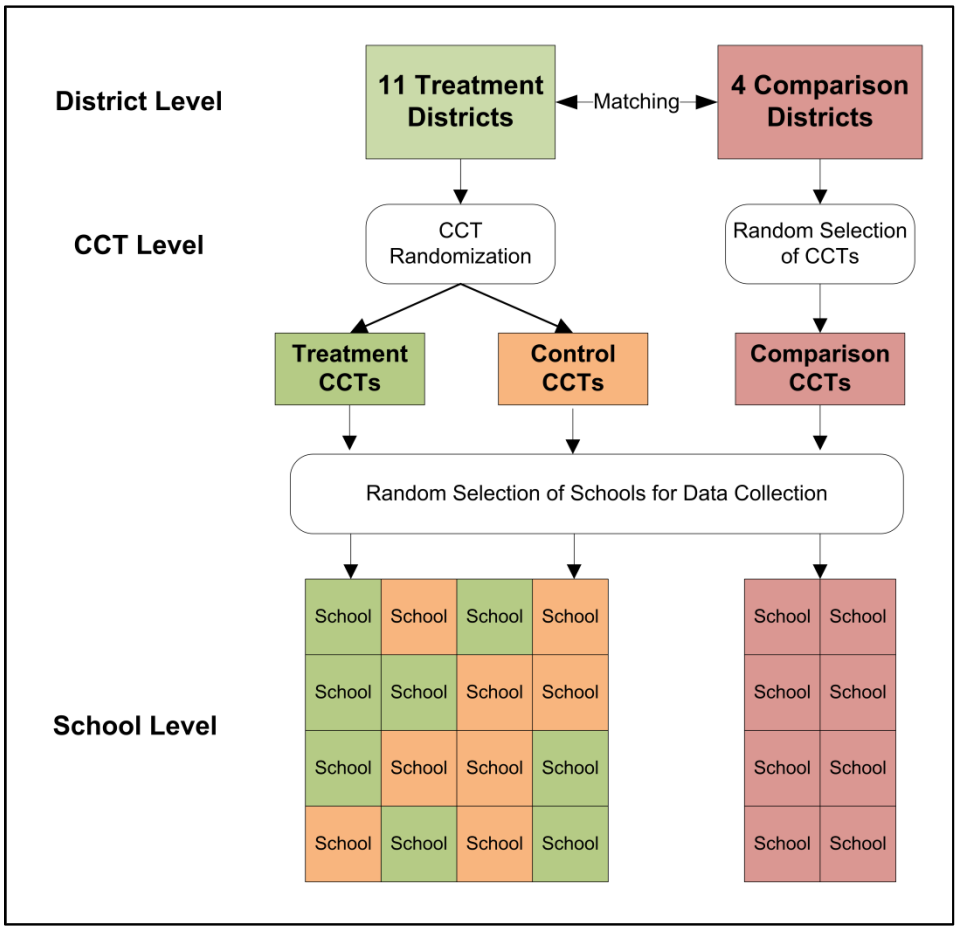
Region	Language Area	Treatment District	Comparison District
Central	Luganda	Gomba Wakiso	Buikwe
East	Ateso	Katakwi Kumi Serere	Ngora
North	Leblango	Apac Lira Kole	Otuke
South West	Runyankore/ Rukiga	Bushenyi Kiruhura Kabale	Ibanda

B.2 ASSIGNMENT OF SCHOOLS TO TREATMENT AND CONTROL GROUPS AND ESTIMATION STRATEGY

The CCTs are responsible for training teachers and providing follow-up support and assistance in implementing the Result 1 interventions. Towards this end, SHRP conducted training workshops (training of trainer workshops) for CCTs in different regions. Since each CCT is responsible for several schools, randomizing at the school level would imply that a CCT would have to treat schools under his or her jurisdiction differently if some were designated as treatment schools and others as controls. Therefore, randomization at the school level had a high risk of 'contamination' between treatment and control groups. To avoid this potential problem, we opted to randomize at the CCT level, assigning the entire cluster of schools under a CCT to either the treatment or the control group. In comparison districts we randomly selected CCTs whose school clusters will serve as non-intervention-district controls.

Figure 1 illustrates the evaluation design, i.e. random assignment into the treatment or control group within treatment districts and random selection of CCTs within comparison districts.

Figure 1. Assignment to Treatment, Control and Comparison Groups at the District and CCT levels



The difference in outcome indicators between treatment schools (green) and control schools (orange) will show the effect of the school-level intervention, given that both types receive the district-level intervention but only treatment schools receive the school-level program. The difference in outcomes between control schools and comparison schools (red) can identify the effect of the district-level intervention. While none of those schools benefits from the school level programs, the control schools are exposed to the district level treatment. The impact of the complete intervention package can be measured by estimating the difference in outcomes between treatment and comparison schools.

We estimate the impact of the first two years of the intervention for each language on a number of literacy outcomes using a difference in differences model. To evaluate the impact of the intervention package on literacy score, *Y*, for students in cluster I within a specific language group, we use the scores collected at the beginning of the school year 2013 (baseline) when they entered P1 and from the end of their second year of school when they finished P2 (endline), and regress the individual test score *Y* of student *i* on the treatment status *T* of the school *s* in district *d,* a dummy variable *P* indicating period (i.e. 0 or 1 respectively for baseline and endline) and their interaction *T***P*:

$$Y_{isdp} = \beta_0 + \beta_1 T_{sd} + \beta_2 P + \beta_3 T_{sd} * P + \beta_4 D_d + \beta_5 X_i + \beta_6 Z_s + \varepsilon_{isd} \quad (1)$$

where X_i are individual characteristics of the student *i*, such as age and sex, Z_s are school characteristics such as language of instruction, D_d is a vector of dummy variables indicating the districts. As we mentioned T_{sd} is a dummy equal to 1 if school *s* in district *d* received the intervention and 0 otherwise, *P* indicates the period –baseline or endline, therefore, our coefficient of interest is $\beta 3$, which shows the differential effect of the treatment at the endline.

B.3 IMPACT INDICATORS AND DATA COLLECTION INSTRUMENTS

Literacy is comprised of multiple skills, both receptive and productive. Successful readers must be able to identify letters and their corresponding sounds, segment and blend those sounds to form words and sentences, master appropriate vocabulary, and make meaning from text, among other skills. They must also be able to demonstrate their understanding and engagement with text through writing. To assess the effectiveness of the SHRP in reaching its goal to improve early grade reading and transition to English, specific key literacy skills were assessed.

The consensus among the reading research community in the United States is that effective reading instruction attends to *at least* five main reading skill areas including alphabetics (letter knowledge and phonemic awareness), fluency, vocabulary, and comprehension and phonics (National Reading Panel, 2000; Snow, Burns & Griffin, 1998).[4] Based on this research, the Early Grade Reading Assessment, (EGRA) a brief oral reading assessment that tests these skills, is used to measure program impacts on literacy (RTI International, 2007).

EGRA is comprised of multiple sub-tests that focus on the five main reading skill areas outlined above. Within each of these five areas, there are multiple sub-tests that can be selected for inclusion, based on local needs and the goals of the assessment system.

No clear benchmarks for the EGRA tool have been established. That is, the EGRA tool provides a snapshot of early literacy skills but does not provide guidelines for interpreting which children can be considered "readers" or what level of performance should be expected on each sub-test. At the same time, EGRA has been used to assess early literacy skills in more than 50 countries around the world; thus, performance of students participating in SHRP can be compared with the range of performance of other children on EGRA in other low-income countries.

A notable component of SHRP is its transitional bilingual design. That is, literacy instruction begins in one of four mother tongue languages, with English introduced as a subject area nearly simultaneously (within 4-8 weeks after mother tongue instruction has begun). The language of instruction will then increasingly transition from mother tongue to English over the course of four years. Because of this transitional bilingual design, the impact evaluation necessarily requires a heteroglossic[5] approach to assessment. Therefore, early literacy skills are assessed in mother tongue and English.

This transitional bilingual design affected the selection of sub-tests included in the EGRA tool in each language. For example, because most grade 1 students cannot be expected to have prior knowledge in English language or literacy, the sub-tests that have been selected to assess literacy in English are aimed at capturing lower skill levels; in contrast, students are expected to already possess basic linguistic knowledge in their mother tongues and the EGRA sub-tests that have been selected aim to capture a distribution of literacy skills that include higher level abilities. Table 2 describes the sub-tests that are included in English and mother tongue EGRA.

[4] These five skills are not meant to be all inclusive; however, considerable empirical research has been conducted in these skill areas that has indicated they are important predictors of reading.

[5] A "heteroglossic" approach conceptualizes literacy learning in both mother tongue and English as interconnected, co-existing, and mutually reinforcing.

Table 2. Early Literacy Skills, EGRA Subtasks

Early Literacy Skill	Sub-test	Measurement	English	Mother -Tongue
Alphabetic Knowledge	Letter Sound Knowledge	Number of letter sounds correctly identified out of 100 in 60 seconds	X	X
Phonemic Awareness	Segmenting	Number of words correctly segmented out of 10 words	X	X
Phonics/Alphabetic Principles	Nonword decoding	Number of nonwords correctly decoded out of 50 in 60 seconds	X	X
Fluency	Oral passage reading	Number of words in a reading passage of approximately 50-60 words read fluently (with accuracy) in 60 seconds	X	X
Reading Comprehension	Oral recall	Number of questions (out of five) about a reading passage (read by student) answered correctly	X	X
Listening Comprehension	Oral recall	Number of questions (out of three) about a passage read aloud (by facilitator) answered correctly	--	X
Receptive Vocabulary	Oral identification of common objects	Number of common objects correctly identified	X	--

In addition to the EGRA, the following data collection instruments were developed and administered:

- Learner Context interview: to determine learner attendance to pre-school, socio-economic status and home literacy environment
- Teacher interview: to determine language of instruction in the school, teacher qualifications (including attendance to SHRP trainings) and the amount of support received from head teachers and CCTs
- Head Teacher interview: to obtain school enrollment information and determine participation in SHRP Training and amount of support provided to teachers
- School inventory: to determine quality of school environment (access to electricity, water, functioning toilets/latrines, availability of a school library)
- Classroom observation: to determine the extent to which teachers are applying SHRP teaching practices

B.4 SAMPLING

The standard approach to determining sample size for analytical surveys is to estimate the sample size required to achieve a specified level of power (probability), such as 90 percent, for detecting a change of a specific magnitude. This sample size depends on a number of factors including the evaluation design, the impact estimate, design of the sample survey used to collect data, the statistical test, and the population under investigation.

The initial group of 11 treatment districts located in the 4 different language areas for Cluster 1 was selected by the IP and MoES to participate in the intervention. NORC selected a sample of 4 comparison districts. Each comparison district was individually matched on the basis of P3 and P6 NAPE literacy scores to each of the 4 language areas. Within each area CCTs were randomly assigned to one of four arms of the intervention (3 treatment arms and 1 control arm). NORC calculated the number of schools needed in each language/arm cell (20 cells) and within each cell, RTI selected the requisite sample of N treatment schools and N control schools using random assignment, and N comparison schools in each comparison district. This "balanced" design is an efficient one, with high return of precision and power for survey resources expended.

For the impact evaluation of the Reading Program, we estimated the sample required to detect a double-difference measure of impact of magnitude D = 0.20 with a power of 90%. These calculations determined that 14 schools were necessary for each of the 20 cells (3 treatment arms, 1 control arm and 1 comparison arm, and 4 language subgroups), and that 30 students would be sampled for each school, for a total of 420 students per cell. With 20 cells, the total sample size required amounted to 8,400 students from 280 schools; i.e. 8,400 students at baseline and 8,400 students at endline. Of the 280 schools, 168 (5,040 students) would constitute the treatment group, and 112 schools (3,360) would be control/comparison schools. In the first year of intervention, 30 P1 students per school would be sampled, and in the second year of intervention 30 P2 students per school would be sampled, given that we are following the same cluster of students over time. Note therefore that the sample design involves a panel of schools but does not involve a panel of students as students are randomly re-selected for every round of data collection.

Based on these estimations, RTI randomly selected some CCTs that included 168 schools for the evaluation sample, from the universe of treated CCTs (410 intervention schools). Control schools within the treatment districts were selected from the schools in those districts that were not selected for the intervention. Table 3 shows the target sample size for the baseline data collection and for the October 2014 follow up data collection

Table 3. Baseline (Round 1) and Follow-up (Round 3) Target Sample Size

Language Group	Treatment Arm 1	Treatment Arm 2	Treatment Arm 3	Control	Comparison	Total
Luganda	14 schools 420 students	14 schools 420 students	14 schools 420 students	14 schools 420 students	14 schools 420 students	70 schools 2100 students
Ateso	14 schools 420 students	14 schools 420 students	14 schools 420 students	14 schools 420 students	14 schools 420 students	70 schools 2100 students
Leblango	14 schools 420 students	14 schools 420 students	14 schools 420 students	14 schools 420 students	14 schools 420 students	70 schools 2100 students
Runyankore /Rukiga	14 schools 420 students	14 schools 420 students	14 schools 420 students	14 schools 420 students	14 schools 420 students	70 schools 2100 students
Total	**168 schools 5040 students**			**56 schools 1680 students**	**56 schools 1680 students**	**280 schools 8400 students**

The sample size for the second follow-up data collection after two years of intervention (Cluster 1 Round 3) matched the baseline sample size, and all schools that were part of the first data collection (baseline) were re-surveyed for the Round 3 data collection.

In October 2013 a smaller randomly selected subsample of learners was assessed. Given that during the first year of the intervention only one treatment was rolled out fewer observations were sufficient. This Round 2 subsample was used to evaluate the impact of the program in its first year (February to October 2013).

C. DATA COLLECTION AND FINAL SAMPLE

Data collection was conducted at the beginning and at the end of the 2013 school year as well as at the end of the 2014 school year by the IP and its local partner Center for Social Research (CSR). The data collection was conducted using tablets in Tangerine software and consisted of EGRAs in English and 4 local languages (each student was assessed in his/her local language and in English), followed by a learner context interview. At each school, field teams also administered a teacher and head teacher interview and a school inventory. Finally classroom observations were also conducted for a subset of schools (~10%).

C.1 ASSESSOR TRAINING[6]

In preparation for each round of data collection, RTI conducted a seven-day training program, and then selected the highest-performing 70 assessors for the data collection. Assessors were selected based on inter-rater reliability tests (IRR) that were given throughout the week as well as interpersonal and leadership skills. Technical training, which was undertaken by School Health and Reading Program staff (RTI and CSR), included hands-on practice, where assessors spent one day in a school administering the tool to learners and teachers. Trainees were first trained to administer the tools on paper and then introduced to electronic data collection on Nexus tablets so that they could be prepared for both circumstances. Four Data Quality Assurance (DQA) Officers, who also acted as assessor trainers, were given an extra day of training on the classroom observation instrument. This included going to a school, observing a class, and comparing findings among one another.

C.2 DATA COLLECTION

Data were collected from February 15 through March 20, 2013 for the baseline round, from October 14 through November 1 2013 for Round 2 and from October 6 to October 24, 2014 for Round 3. Approximately seventy assessors in teams of four (each of these teams included one supervisor) were deployed to the four language areas. Each of the four language area teams was supported by a DQA Officer. This DQA Officer was responsible for overseeing all aspects of data collection deployment, observing assessors and providing feedback and support, ensuring data were uploaded from the electronic tablets every evening, and observing the reading classes. Besides the team supervisors and the DQA Officers, data collection was overseen by MoES staff (from the Directorate of Education Standards and from Guidance and Counseling), Uganda National Examination Board (UNEB) staff, School Health and Reading Program Staff, and staff from the external evaluation team from NORC.

C.3 SCHOOLS AND LEARNERS SAMPLED

The final sample included in the impact analysis after two years of intervention consisted of a panel of 273 schools and included 7,273 students at baseline (86.6% of the target of 8,400 students) and 7,245 students in Round 3 (86.25% of the target of 8,400 students). The numbers are lower than the target but we expect that it would be still enough to conduct a rigorous analysis. Table 4 presents the number of students included in the analysis; the number of schools is indicated in parentheses.

[6] From RTI EGRA Baseline Report

Table 4. Final Sample Used in the Impact Analysis after Two Years of Intervention

	Treatment 1		Treatment 2		Treatment 3		Control		Comparison	
	Baseline	**Round 3**	**Baseline**	**Round 3**	**Baseline**	**Round 3**	**Baseline**	**Round 3**	**Baseline**	**Round 3**
Luganda	353	304	369	379	321	335	306	343	236	251
	(14)		(14)		(14)		(14)		(12)	
Lango	420	396	421	379	417	393	364	359	388	346
	(14)		(14)		(14)		(12)		(13)	
Ateso	422	423	425	379	422	404	394	362	418	420
	(14)		(14)		(14)		(13)		(14)	
Runyankore Rukiga	311	334	346	361	314	326	289	364	337	387
	(14)		(14)		(14)		(13)		(14)	
Total	**1,506**	**1,457**	**1,561**	**1,498**	**1,474**	**1,458**	**1,353**	**1,428**	**1,379**	**1,404**
	(52)		**(52)**		**(52)**		**(48)**		**(49)**	

D. BALANCE AT BASELINE

In order to explore comparability of pre-treatment characteristics of the treatment, control and comparison groups, we conducted mean equality tests to test for balance. We compared pre-treatment scores (1) between treatment and control schools within treatment districts and (2) between treatment schools from treatment districts and comparison schools from comparison districts. Because comparisons between languages cannot be made given that no benchmarking and psychometric analysis has been performed to assess comparability between languages, mean equality tests of local language literacy scores are conducted within language subgroup. On the other hand, tests for English literacy scores are conducted on the entire sample.

D.1 Treatment vs. Control Schools within Treatment Districts

Overall, our analysis indicates that students' characteristics between treatment and control groups are balanced. Table 5 shows that the treatment and control groups are very similar in terms of demographics and other socio-economic status (SES) characteristics at baseline. Most of the variables included in the means equality tests are not statistically significantly different between treatment and control groups. We find a small difference in the number of female teachers, where the treatment group shows a higher proportion of women although the difference is only borderline statistically significant (p-value=0.08). It is not unusual that among so many variables we would find that one of the tests indicates a difference. Although these socio-demographic variables are not statistically significantly different between the two groups, we will include them as covariates in our analysis of the impact of SHRP on the outcomes of interest.

Table 5. Demographics and other characteristics at baseline for treatment and control groups. All language groups.

Variable	Treatment Mean (SE)	Control Mean (SE)	Difference (C – T)
Age of student	7.338 (0.028)	7.334 (0.053)	-0.004
% of cases with missing age information	0.156 (0.005)	0.152 (0.010)	-0.004
Gender of student (female = 1)	0.492 (0.007)	0.494 (0.013)	0.002
Number of assets (max = 8)	2.513 (0.018)	2.533 (0.032)	0.020
Lives with both parents (yes = 1)	0.633 (0.007)	0.668 (0.013)	0.035
Does not live with mother (yes = 1)	0.177 (0.006)	0.173 (0.010)	-0.004
Reads at home (yes = 1)	0.422 (0.007)	0.417 (0.013)	-0.005
Attended preschool (yes = 1)	0.461 (0.007)	0.416 (0.013)	-0.045
Student absent any day in the week prior to assessment (yes = 1)	0.523 (0.007)	0.508 (0.014)	-0.015
Teacher absent any day in the week prior to assessment (yes = 1)	0.433 (0.008)	0.435 (0.014)	0.002
Gender of teacher (female = 1)	0.756 (0.006)	0.615 (0.013)	-0.141*

Note: * Significant at 10% level

We also compare the basic literacy subtasks scores in English and in local language. We focus on the 3 most basic reading subtasks given that at baseline most of the students are unable to complete any of the more complex items and therefore those scores are noisy and uninformative. Table 7 shows the comparison by language for the letter sound, word segmenting and non-word decoding scores. There are some small differences between treatment and control groups in some of the scores for some languages but none of them are statistically significant. In any case, our analysis will control for the initial levels of competence of the students at baseline.

Table 6. Literacy scores, basic subtasks, by language, for treatment and control groups

Literacy Scores	Treatment Mean (SE)	Control Mean (SE)	Difference (C – T)
English			
Letter sound score (max = 100)	1.543 (0.059)	1.522 (0.092)	-0.021
Word segmenting score (max = 10)	0.221 (0.014)	0.188 (0.024)	-0.033
Nonword decoding score (max = 50)	0.140 (0.018)	0.085 (0.026)	-0.055
Runyankore/Rukiga			
Letter sound score (max = 100)	3.070 (0.171)	3.388 (0.271)	-0.319
Word segmenting score (max = 10)	3.278 (0.131)	3.584 (0.229)	-0.306
Nonword decoding score (max = 50)	0.423 (0.018)	0.406 (0.026)	0.017
Luganda			
Letter sound score (max = 100)	2.419 (0.157)	1.852 (0.259)	0.568
Word segmenting score (max = 10)	2.351 (0.115)	3.129 (0.272)	-0.777
Nonword decoding score (max = 50)	0.199 (0.044)	0.140 (0.078)	0.060
Lango			
Letter sound score (max = 100)	0.785 (0.085)	0.693 (0.160)	0.092
Word segmenting score (max = 10)	0.020 (0.008)	0.005 (0.004)	0.015
Nonword decoding score (max = 50)	0.016 (0.012)	0.015 (0.015)	0.000
Ateso			
Letter sound score (max = 100)	1.810 (0.106)	1.782 (0.196)	0.028
Word segmenting score (max = 10)	1.725 (0.094)	1.380 (0.147)	0.344
Nonword decoding score (max = 50)	0.015 (0.008)	0.067 (0.041)	-0.052

D.2 Treatment Schools from Treatment Districts vs. Control Schools from Comparison Districts

By contrast, the baseline scores between treatment and comparison groups (i.e. treatment schools within treatment districts and control schools within comparison districts) were not as well balanced. This was expected given that comparison districts were selected through matching, rather than randomization and the information to construct the matching was limited. It is also probably the result of reducing the number of comparison districts to only 4, rather than the 11 planned in the original design. Most of the differences between these two groups were not statistically significant however there are imbalances, particularly in some basic literacy subtasks

While there were no significant differences between treatment and comparison groups in basic demographic characteristics from the entire sample, we do note significant differences between treatment and comparison groups for two variables: (1) the number of assets is higher among treatment students and (2) a higher proportion of treatment students had attended preschool than comparison students (Table 7).

Table 7. English literacy scores, demographics and other characteristics at baseline for treatment and comparison groups. All language groups

Variable	Treatment Mean (SE)	Comparison Mean (SE)	Difference (C – T)
Age of student	7.338 (0.028)	7.334 (0.050)	-0.004
% of cases with missing age information	0.156 (0.005)	0.152 (0.010)	-0.004
Gender of student (female = 1)	0.492 (0.007)	0.492 (0.014)	0.000
Number of assets (max = 8)	2.513 (0.018)	2.325 (0.034)	-0.188***
Lives with both parents (yes = 1)	0.633 (0.007)	0.658 (0.013)	0.025
Does not live with mother (yes = 1)	0.177 (0.006)	0.165 (0.010)	-0.012
Reads at home (yes = 1)	0.422 (0.007)	0.417 (0.014)	-0.005
Attended preschool (yes = 1)	0.461 (0.007)	0.395 (0.013)	-0.066*
Student absent any day in the week prior to assessment (yes = 1)	0.523 (0.007)	0.533 (0.014)	0.010
Teacher absent any day in the week prior to assessment (yes = 1)	0.433 (0.008)	0.471 (0.014)	0.038
Gender of teacher (female = 1)	0.756 (0.006)	0.655 (0.013)	-0.101

Note: * Significant at 10% level, ** Significant at 5% level

Within each language subgroup as well, literacy scores are not as well balanced between treatment and comparison groups as they were between treatment and control groups (Table 8). We find significant differences between treatment and comparison groups in the following variables and for the following language subgroups: (1) R/R, Luganda, and Ateso treatment students scored statistically significantly better than comparison students in letter sound identification and (2) R/R treatment students scored statistically significantly better than comparison students in nonword decoding.

Table 8. Literacy scores, basic subtasks, by language, for treatment and comparison groups

Literacy Scores	Treatment Mean (SE)	Comparison Mean (SE)	Difference (C – T)	
Runyankore/Rukiga				
Letter sound score (max = 100)	3.070 (0.171)	1.612 (0.219)	1.458	**
Word segmenting score (max = 10)	3.278 (0.132)	3.609 (0.254)	-0.331	
Nonword decoding score (max = 50)	0.423 (0.070)	0.000 (0.000)	0.423	**
Luganda				
Letter sound score (max = 100)	2.420 (0.157)	1.454 (0.204)	0.966	**
Word segmenting score (max = 10)	2.351 (0.115)	2.467 (0.227)	-0.115	
Nonword decoding score (max = 50)	0.200 (0.044)	0.176 (0.076)	0.023	
Lango				
Letter sound score (max = 100)	0.785 (0.086)	0.885 (0.160)	-0.099	
Word segmenting score (max = 10)	0.021 (0.008)	0.003 (0.003)	0.018	
Nonword decoding score (max = 50)	0.016 (0.012)	0.027 (0.020)	-0.012	
Ateso				
Letter sound score (max = 100)	1.810 (0.106)	1.365 (0.186)	0.445	*
Word segmenting score (max = 10)	1.725 (0.094)	1.551 (0.171)	0.174	
Nonword decoding score (max = 50)	0.015 (0.008)	0.000 (0.000)	0.015	

Note: ** Significant at 5% level, * Significant at 10% level.

Overall, the sample at baseline was well-balanced between treatment and control schools indicating that the randomization generated equivalent groups in terms of baseline characteristics. The sample is less well-balanced between treatment and comparison schools. In our analysis we take advantage of the fact that we have baseline data and we control for original assessment scores and for learners characteristics. The differences found between treatment and comparison schools will be taken into account therefore in the data analysis but may imply some limitations in the comparison.

E. IMPACT ANALYSIS RESULTS

This section summarizes the results of the impact analysis after two years of intervention for Cluster 1 students who were in first grade during the first year of intervention and then moved on to second grade during the second year of intervention. Our analysis is conducted for each of the outcome scores on subtasks conducted in the local language for each of the four local languages. On the other hand, for outcome scores on English subtasks, our analysis is applied to the entire sample (pooled sample from all four language subgroups). It is important to keep in mind that this evaluation does not allow for comparison of progress and impact between languages as the EGRA tools were developed independently for each local language and no psychometric analysis has been conducted to determine whether EGRA scores in one language can be compared to EGRA scores in another language. Therefore, while we can determine whether SHRP has had an impact for a particular language subgroup vs. another language subgroup, we cannot comment on the relative magnitude of these impacts between language subgroups.

E.1 IMPACT OF SCHOOL-LEVEL INTERVENTION: TREATMENT VS. CONTROL SCHOOLS

First, we study the effect of the school-level intervention following the model in equation (1) using the treatment and control school sample (see Section B.2, Figure 1). We present findings for scores in local language within each language subgroup and in English for the pooled sample. In Tables 9 to 13 below, we report the average treatment effect on each outcome of interest and for each language subgroup using different model specifications to check for robustness. In general, the regression models include individual controls (sex and age of the student, a dummy for whether age is missing, household asset index, dummies for living with both parents, having someone read to the student at home, and language spoken at home)[7] and district dummies. For the regressions on English scores using the pooled sample, we include dummies for local language subgroups. We also test results with and without school fixed effects. All standard errors are robust standard errors and allow for correlation in the unobservables between learners in the same class. All regressions are Ordinary Least Squares (OLS) regressions.

Impact on Letter Sound Knowledge

Each cell in Table 9 shows the average effect of the treatment (SHRP intervention) on the Letter Sound Knowledge Score for each language subgroup using 5 different model specifications. Since schools were randomly assigned to the treatment and control groups, and mean equality tests show that both groups are pretty well balanced at baseline, we start with the simplest model (model (1)) which does not include any controls. To reach more precision and to check for robustness, we add controls in models (2) through (5).

- Model (1) includes no controls.
- Model (2) is similar to model (1) but includes districts fixed effects.
- Model (3) is similar to model (2) but adds individual controls.
- Model (4) is similar to model (3) and also includes school fixed effects.

- Model (5) is similar to model (4) but only includes the subsample of students 6 years and older. We do this because a somewhat considerable proportion of students interviewed at baseline and endline reported being 4 and 5 years of age even though the official age for starting P1 in Uganda is 6 years old. At baseline, 4.5 percent of the students in our sample reported being 4 and 5 years of age. Although it may be possible that some students officially enroll in P1 before the age of 6, reports from the data collection teams suggest that these younger children present in P1 classes are in fact not formally enrolled in primary school. This issue was uncovered at the beginning of the baseline data collection when teams were already in the field; therefore, the issue of whether to include them in the data collection was not addressed during training (since it was assumed that all students present would be official P1 students). It is, therefore, possible that field teams sampled children who were too young to attend P1. In the follow-up round, in addition to the natural aging process of the cluster, this issue was handled in a more systematic way (i.e. excluding those children that did not formally belong to the P1 classes). For these reasons, for this analysis, we decided to include a model that only analyzes the subsample of students who reported being 6 years and older, i.e. students more likely to be officially enrolled in P1.

With the simplest model (model (1)), we find that, after two years, the intervention had a positive impact on Letter Sound scores in Luganda, Lango and English. We find that results are very similar across all other models; the intervention, after two years, shows positive effects for Luganda, Lango, and English, while for Runyankore/Rukiga and Ateso the effect, although positive, is not statistically significantly different from zero.

[7] With the exception of models (1) and (2) in

Table 9. .

Table 9. SHRP School Level Effect, Year 2 - Letter Sound Score

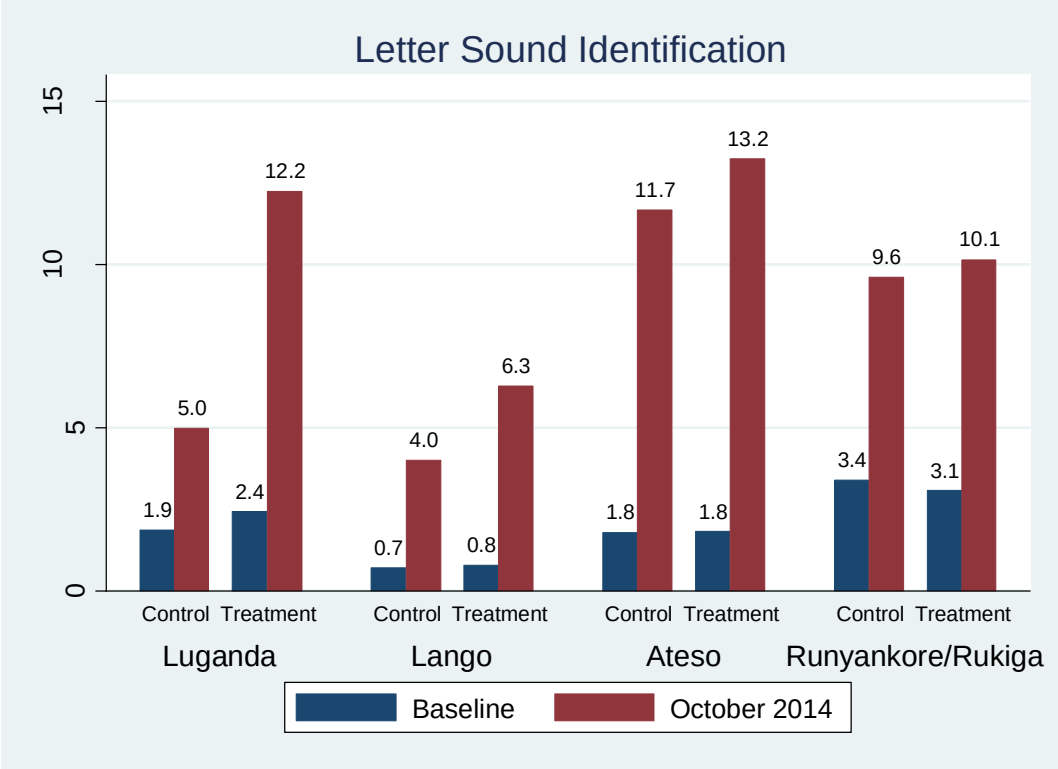
	(1)	(2)	(3)	(4)	(5)
Language					6 years and older
Runyankore/Rukiga	0.855	0.848	0.589	0.906	0.830
	(1.168)	(1.158)	(1.157)	(1.123)	(1.173)
Luganda	6.686***	6.865***	6.920***	6.732***	6.949***
	(1.060)	(1.035)	(1.048)	(1.024)	(1.003)
Lango	2.194***	2.202***	2.265***	2.246***	2.183***
	(0.718)	(0.714)	(0.730)	(0.724)	(0.737)
Ateso	1.539	1.505	1.712	1.661	1.787
	(1.429)	(1.425)	(1.366)	(1.380)	(1.336)
English	2.298***	2.382***	2.343***	2.345***	2.434***
	(0.628)	(0.638)	(0.614)	(0.606)	(0.597)
Districts Fixed Effects	no	yes	yes	yes	yes
Individual Controls	no	no	yes	yes	yes
School Fixed Effects	no	no	no	yes	yes

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Individual controls include age and sex of learner, dummy for age missing, score on household assets index, dummies for student living with both parents, someone at home reading to the student and language spoken at home. Regressions for English scores include controls for local language.

The larger positive effect is seen in the Luganda subgroup where the treatment students scored on average 6.8 letters sounds (0.59 standard deviation) higher than control students due to the program. The program effect among the Lango subgroup is also positive but more modest. On average, treatment students were able to correctly identify around 2.2 additional letter sounds (0.27 standard deviation) compared to control students as a result of the program.

The raw data, although it cannot reflect the complexity of the econometric analysis shown above, are useful for visualizing the level of knowledge of the students in the different subgroups at baseline and after two years of intervention. Figure 2 shows the average number of correctly identified letter sounds by treatment and control groups by language.

Figure 2: Letter Sound Scores at baseline and after two years of intervention, by language

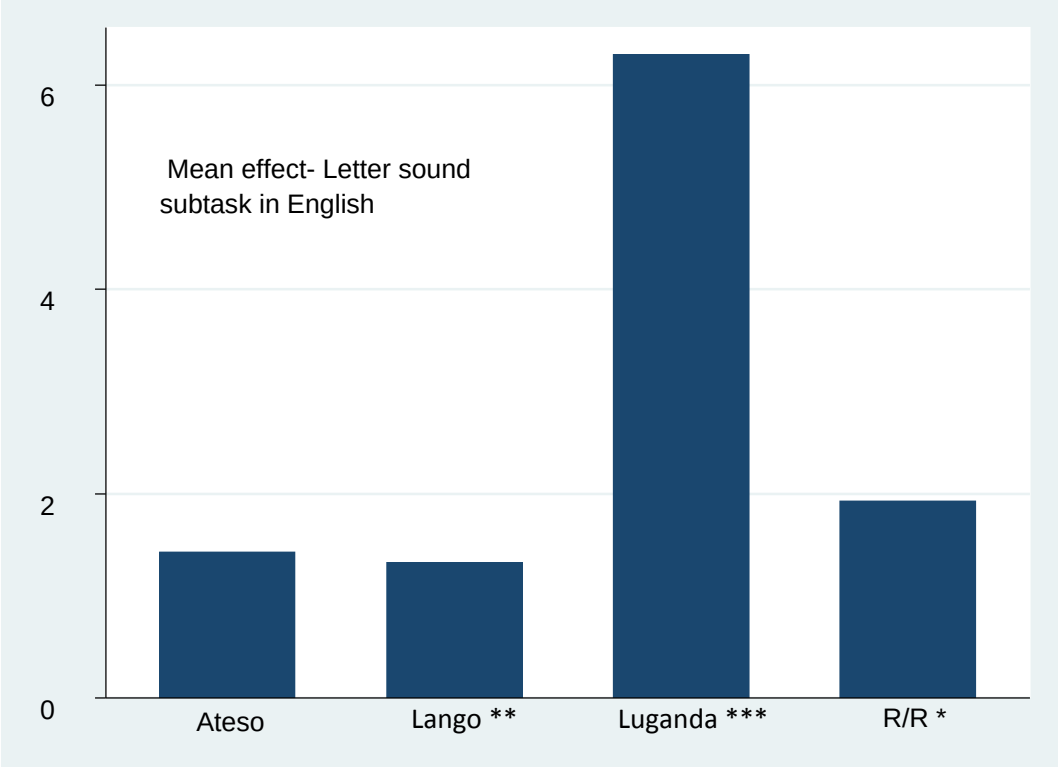


It is encouraging that the program helps students achieve higher scores in Lango, and in Luganda in particular. However, it is important to note that the absolute levels of competence remain low. After two years of intervention, students could only read 12.2 letter sounds per minute in the Luganda treatment sub-group, and 6.3 letter sounds per minute in the Lango treatment sub-group. While the effect size of the improvement is large in Luganda, the absolute scores achieved are low for a student who has completed P2. Furthermore, for the Lango sub-group, the impact in absolute terms only corresponds to a 2.2 letter sound increase as compared to the control group.

In the two languages where we find no significant effect of the program we observe that at the end of P2, learners in the treatment group scored, on average, 13.2 letter sounds per minute in the Ateso sub-group and 10.1 letter sounds per minute in the R/R treatment sub-group. The performance of the control learners is slightly lower but not statistically different. The average score among control group students is 11.7 letter sounds among the Ateso learners and 9.6 among the R/R group.

The impact of the intervention on the English version of this subtask is approximately 2.3 letter sounds (Table 9, Row 5); that is, on average, treatment students were able to recognize 2.3 more English letters (0.25 standard deviation) than their counterparts in control schools. Most of this effect comes from a differential impact among the students in the Luganda sub-group. Figure 3 shows the mean effect of the program on the letter sound subtask in English for each sub-language. While all language sub-groups show some positive effect, the improvement in Ateso, Lango and R/R is less than a 2 letter sound increase and in some instances it is not significant, as is clearly the case for the Ateso sub-group. As we mentioned, the Luganda language is the sub-group that shows a large increase in the English version of the task.

Figure 3: Mean effect of the program on Letter Sound Scores in English, by language group



*** p<0.01, ** p<0.05, * p<0.10

In general, students that perform well in the letter sound identification subtask in English are the ones that perform well in the local language version of the subtask. However this is more pronounced for the Luganda subgroup, where the correlation between the 2 language versions of the subtasks reaches 0.80 while for other languages it is around 0.60-0.70.

Table 10 shows the average treatment effect on the percentage of students who could not correctly identify any letters in the letter sound knowledge subtask (percentage of zero scores). Columns (1) and (2) show the results of regressions without and with school fixed effects respectively, although results are extremely similar. The SHRP intervention had a significant impact in reducing the percentage of students who scored zero in Luganda and Lango letter sound knowledge; in both cases, the intervention led to approximately a 17 percentage point (0.38 standard deviation) decrease in the proportion of students who could not identify any letters. We also find a significant impact on the percentage of students who scored zero on the English letter sound knowledge subtask; the intervention contributed to a 13 percentage point (0.29 standard deviation) drop in the students who knew no English letter sounds. Again, most of this effect is due to the contribution of the Luganda speaking students.

In contrast, the program did not produce a significant reduction in the proportion of learners with zero scores for any of other two local languages: R/R and Ateso.

Table 10. SHRP School Level Effect - Percentage of students who scored zero on letter sound subtask

Language	(1)	(2) With school fixed effects
R/R	-0.037	-0.043
	(0.056)	(0.055)
Luganda	-0.174***	-0.172***
	(0.054)	(0.053)
Lango	-0.174***	-0.174***
	(0.046)	(0.046)
Ateso	-0.036	-0.035
	(0.052)	(0.053)
English	-0.127***	-0.132***
	(0.034)	(0.034)

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All regressions include district fixed effects, individual controls and a constant term. Individual controls include age and sex of learner, dummy for age missing, score on household assets index, dummies for student living with both parents, someone at home reading to the student and language spoken at home. Regressions for English language include control for local language.

Impact on Word Segmenting Score

Table 11 shows the average treatment effect on Word Segmenting scores. Students received one point for every word that they could segment correctly for a maximum of 10 points. We find a positive and significant effect for the Luganda subgroup for which the treatment students scored on average approximately 1.5 words (0.5 standard deviation) better than the control students. In contrast the Lango subgroup shows a significant and negative effect of the program; the average treatment student scored around 0.46 fewer words (0.29 standard deviation) than the average control student, which is a surprising finding. No impact is found for any other local languages, nor in English word segmenting.

Table 11. SHRP School Level Effect - Word Segmenting Score

Language	(1)	(2) With school fixed effects
R/R	0.563	0.527
	(0.515)	(0.512)
Luganda	1.527***	1.527***
	(0.488)	(0.496)
Lango	-0.462***	-0.464***
	(0.150)	(0.148)
Ateso	0.101	0.028
	(0.522)	(0.453)
English	-0.141	-0.144
	(0.157)	(0.157)

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All regressions include district fixed effects, individual controls and a constant term. Individual controls include age and sex of learner, dummy for age missing, score on household assets index, dummies for student living with both parents, someone at home reading to the student and language spoken at home. Regressions for English language include control for local language.

Impact on Non-Word Decoding, Oral Reading Fluency, Reading Comprehension and English Receptive Vocabulary

In terms of the impact of the intervention on higher level literacy skills, namely non-word decoding, oral reading fluency, reading comprehension and English receptive vocabulary, we find small positive effects. Again, Luganda displays larger positive effects due to the program, while the effects for Lango and Ateso are smaller and sometimes only borderline significant (i.e. non-word decoding for the Lango subgroup). In Table 12 we show our findings using the school fixed effect models[8].

Table 12. SHRP School Level Effect - Non-Word Decoding, Oral Reading Fluency, Reading Comprehension and English Receptive Vocabulary

Dependent Variable Language	(1) Non-Word Decoding	(2) Oral Reading Fluency	(3) Reading Comprehension	(4) English Receptive Vocabulary
R/R	2.272** (0.954)	3.085*** (0.925)	0.219** (0.091)	
Luganda	2.761*** (0.940)	3.508** (1.372)	0.263** (0.115)	
Lango	0.492* (0.276)	1.342** (0.532)	0.101*** (0.029)	
Ateso	0.849** (0.374)	0.980*** (0.314)	0.107*** (0.033)	
English	1.197*** (0.396)	1.654** (0.682)	0.086** (0.036)	0.143 (0.209)

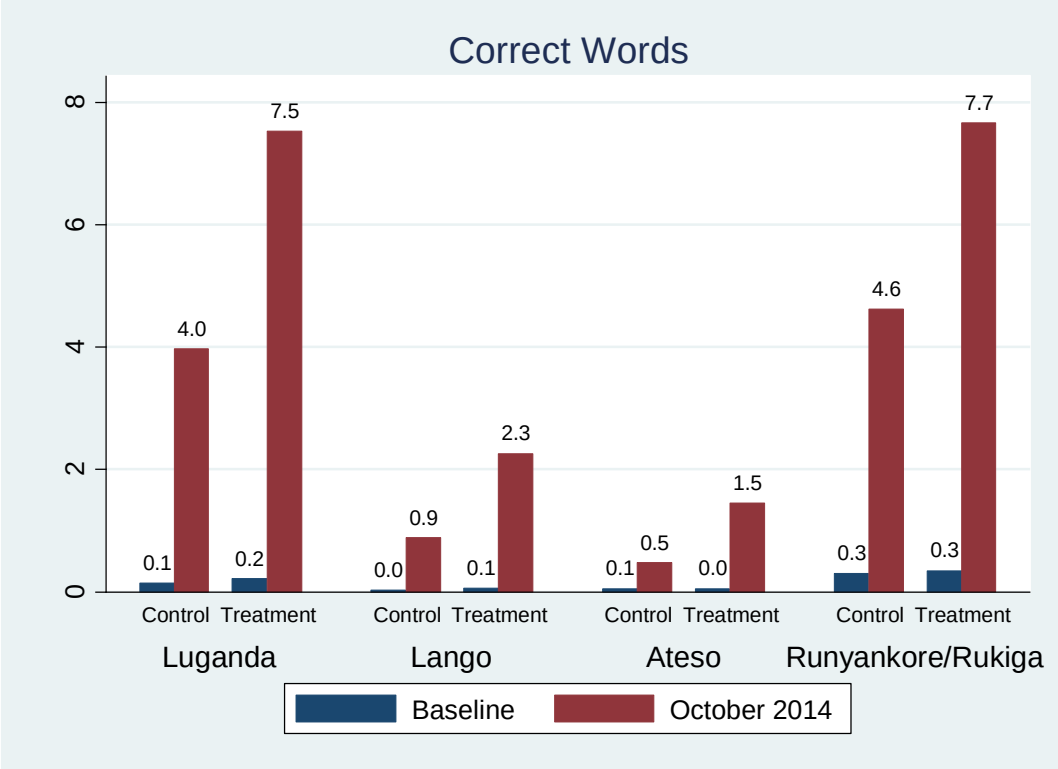
Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All regressions include district and school fixed effects, individual controls and a constant term. Individual controls include age and sex of learner, dummy for age missing, score on household assets index, dummies for student living with both parents, someone at home reading to the student and language spoken at home. Regressions for English language include controls for local language.

The program increased the number of non-words that students can properly decode by 2.3 (0.27 standard deviation) and 2.8 words (0.33 standard deviation) in the case of R/R and Luganda respectively. The impact for the Ateso language sub-group is smaller at 0.8 words (0.19 standard deviation), while for Lango it is an average 0.5 words (0.15 standard deviation) and only statistically significant at the 10% level. Similarly the effects of the program on reading fluency and comprehension are more pronounced in Luganda and R/R than in Lango and Ateso where they are quite small but positive and significant nevertheless. Again, it is encouraging that the intervention has positive effects in most of these tasks, although absolute levels of achievement remain low.

[8] Results are very similar when we do not include school fixed effects.

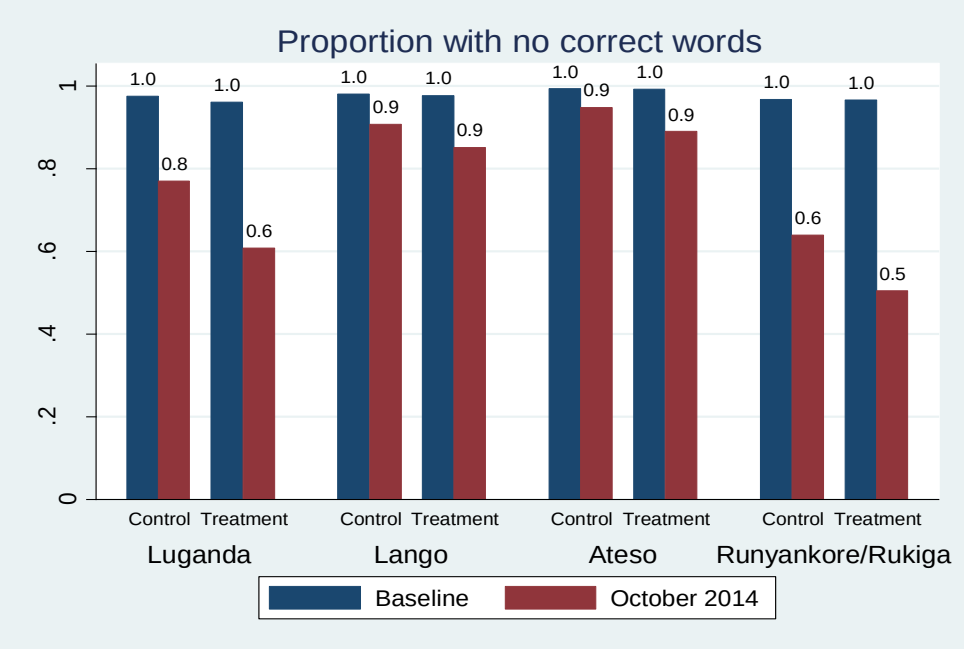
Figure 4 shows the average number of words that students in each sub-group were able to read correctly in a reading passage in 60 seconds at baseline and after two years of intervention. The numbers are low in general and particularly for Lango and Ateso languages where scores in other reading competencies are also low. For example, treatment P2 students in the Lango and Ateso subgroups were, on average, only able to read 2.3 and 1.5 words per minute respectively, even after 2 years of program intervention. Even in the Luganda and R/R subgroups, where program impact on reading fluency was more pronounced, students in the treatment group were only able to read 7.5 words and 7.7 words per minute respectively.

Figure 4: Oral reading fluency at baseline and after two years of intervention, by language.



Even more revealing is the fact that there is a large number of learners that were not able to read a single word correctly from the reading passage. This can be seen in Figure 5 below. Even in the sub-group that performs best - R/R students in treatment schools - half of the treatment group learners are not able to read a single word correctly from the reading passage by the end of second grade. In the less well-performing language groups (Ateso and Lango) the equivalent proportion of treatment students with a zero score in oral reading fluency is 90 percent.

Figure 5. Proportion of students who scored zero on oral reading fluency



Not only do we find that a large proportion of learners cannot read a single word but we also see that almost none of them can read more than 30 words and very few students are able to read more than 20 words correctly. Table 13 shows the percentage of learners that achieve those levels of oral reading fluency by language and by treatment group in October 2014. The percentage of those that are able to read more than 20 words correctly in general favor the treatment group, reflecting the effect of the program, although the levels are very small, particularly for Lango and Ateso languages. An insignificant proportion –less than 1% in most cases- of students managed to read more than 30 words correctly in any of the languages or treatment subgroups.

Table 13. Oral Reading Fluency, More Than 20 and 30 Words, October 2014

	More than 20 correct words, %		More than 30 correct words, %	
Language	**Treatment**	**Control**	**Treatment**	**Control**
R/R	14.4	4.9	0.1	0.3
Luganda	17.3	7.6	1.3	2.0
Lango	3.7	0.3	0.7	0.0
Ateso	1.7	0.3	0.1	0.0

Impact on Boys vs. Girls

In general we find that girls tend to perform better than boys in reading subtasks with the exception of the segmenting subtask where they perform similarly. This is true for all languages except Ateso where girls only have a very small advantage in letter sound identification scores but not in other subtasks. It is possible that this is the case because, in general, the scores for the Ateso subgroup are very low and many learners are not able to complete the subtasks at all. We do not observe differences between boys and girls in the English receptive vocabulary subtask.

Finally, no differences in the impact of SHRP by gender were identified.

E. 2 IMPACT OF DIFFERENT TREATMENT ARMS

As we mentioned, during the second year of implementation three slightly different interventions, or "treatment arms" were rolled out: (1) Basic Program: teacher trainings alone; (2) Basic Program + manpower support: teacher trainings with a certain number of CCT visits to schools per year; and (3) Basic Program + SMS support: teacher trainings with SMS support by CCTs. All arms included the distribution of textbooks, parental and community awareness, head teachers leadership trainings, etc.

We present results for the language subgroup for which we have systematically found the largest effects of the program, Luganda, to study the differential effects of the three arms. We do this because given the larger average effect in this language, we are more likely to be able to estimate differences between arms with the same precision.

Table 14 shows that, in general, the treatment arm that presumably offers teachers the most support – treatment 3 – has larger effects than the intermediate support arm treatment 2. The basic treatment arm which provides teacher training but not CCT visits or SMS support does not show a significant effect, in most subtasks. The exception to this is the word segmenting subtask, where treatment 1 has a positive effect that is even larger than the two other treatment arms.

It is not totally clear however that the differences between treatment 2 and treatment 3 are always statistically significant. It is likely that in a larger sample or over time, if effects continue to increase, we could detect the differences with more precision. According to the USAID/Uganda SHRP Annual Report October 2013-September 2014, implementation fidelity of the three treatment arms was low. In particular, the implementing agency reports that there were issues with the SMS system and that teachers only received 14 messages during the school year, many of these messages were about the training rather than technical content. The number of teachers that actually received messages is also not known. Furthermore, teachers did not receive the extra support as planned due to the perception that the fuel allowance provided by SHRP was too low[9]. This may explain the inconsistency in some findings displayed in Table 14.

[9] USAID/Uganda School Health and Reading Program, Annual Report, October 2013-September 2014, p.34.

Table 14. SHRP School Level Effect - Non-Word Decoding, Oral Reading Fluency, Reading Comprehension

	(1) Letter Sound	(2) Segmenting	(3) Non-Word Decoding	(4) Oral Reading Fluency	(5) Reading Comprehension
Treatment 1	3.156	2.051***	-0.392	-0.797	-0.064
	(1.980)	(0.528)	(1.117)	(1.462)	(0.124)
Treatment 2	6.879***	1.203**	3.762***	4.765***	0.340**
	(1.640)	(0.545)	(1.064)	(1.564)	(0.133)
Treatment 3	8.506***	1.398**	4.588***	6.120***	0.484***
	(2.017)	(0.577)	(1.055)	(1.643)	(0.135)

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Individual controls include age and sex of learner, dummy for age missing, score on household assets index, dummies for student living with both parents, someone at home reading to the student and language spoken at home. All regressions

E.3 IMPACT OF DISTRICT AND SCHOOL-LEVEL INTERVENTION: TREATMENT VS. COMPARISON SCHOOLS

In addition to assessing the effect of the school-level interventions, we also assess the combined effect of the district and school-level interventions by comparing treatment schools in program districts to comparison schools in non-program districts (see Evaluation Design section, Figure 1). The regressions in Table 15 below all include school and district fixed effects, and individual controls. The results from regression models without these fixed effects and controls are very similar. In addition, we also estimated all these models using propensity-score matching (PSM) at the school level to further ensure comparability of treatment and comparison cases. These models yield similar results to the ones without use of school-level PSM.

Similar to the impact of the school-level intervention discussed in Section E1, we find impacts of the school+district-level intervention on the Luganda and Lango subgroups in letter sound scores. On average, treatment students in the Luganda subgroup could read 6.7 letters more than comparison students while the impact among the Lango students is around 2.4 letters. We do not find much effect of the program on word segmenting, the effects are in general positive but mostly statistically insignificant. In the case of the Lango language sub-group, we observe a negative impact of half a word.

For all languages, our analysis shows that the school+district-level intervention had, in general, small but positive impacts on non-word decoding, oral reading fluency, and reading comprehension. The results are significant for English, Luganda and Lango languages but not for R/R or Ateso. In the case of Ateso the impact of the program on reading comprehension is very small but statistically significant.

In interpreting these findings, it is important to remember that, as noted in Section B1, we were only able to select one comparison district for each local language, due to budget and logistical constraints, resulting in only 4 comparison districts, while there are 11 treatment districts. This limitation in the evaluation design means that the district-level analysis is not as robust and reliable as it would have been, had we been able to select one comparison district for each treatment district, as was originally intended.

Table 15. SHRP School and District Level Effect – Letter Sound Knowledge, Word Segmenting, Non-Word Decoding, Oral Reading Fluency, Reading Comprehension and English Receptive Vocabulary

Language	Letter Sound Knowledge	Word Segmenting	Non-Word Decoding	Oral Reading Fluency	Reading Comprehension	English Receptive Vocabulary
R/R	-0.671	0.548	1.148	2.196*	0.157	
	(0.777)	(0.407)	(1.071)	(1.265)	(0.102)	
Luganda	6.737***	0.933*	2.608***	3.804***	0.308***	
	(1.020)	(0.494)	(0.796)	(1.005)	(0.075)	
Lango	2.441**	-0.553***	0.519**	1.611***	0.077**	
	(1.030)	(0.142)	(0.233)	(0.491)	(0.032)	
Ateso	1.817	-0.107	0.349	0.505	0.086**	
	(1.519)	(0.627)	(0.467)	(0.487)	(0.033)	
English	2.384***	-0.215	1.354***	2.093***	0.116***	-0.201
	(0.609)	(0.136)	(0.339)	(0.581)	(0.031)	(0.201)

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All regressions include district and school fixed effects, individual controls and a constant term. Individual controls include age and sex of learner, dummy for age missing, score on household assets index, dummies for student living with both parents, someone at home reading to the student and language spoken at home. Regressions for English language include control for local language.

Impact on Boys vs Girls

In terms of differential impacts on boys vs girls, the findings are similar to those of the impact of the school-level intervention. We do not find that the program has different impacts depending on the gender of the learner. In addition, in general, girls show an advantage over boys on most subtasks.

SUMMARY OF FINDINGS AND CONCLUSION

This report presents the results of the impact evaluation of the School Health and Reading Program (SHRP) after two years of program interventions designed to improve Early Grade Reading.

Overall, the analyses suggest that, after its first two years of implementation, the program has had a positive impact, particularly in the Luganda language. Table 16 shows a summary of results for all subtasks as well as effect sizes expressed in standard deviations from the school-level analysis which uses an experimental design.

Table 16. SHRP School-Level Effect – Summary Table

Language	Letter Sound Knowledge	Word Segmenting	Non-Word Decoding	Oral Reading Fluency	Reading Comprehension	English Receptive Vocabulary
R/R	0.906	0.527	2.272**	3.085***	0.219**	
	(1.123)	(0.512)	(0.954)	(0.925)	(0.091)	
Effect Size (SD)	--	--	*0.27*	*0.33*	*0.26*	
Luganda	6.732***	1.527***	2.761***	3.508**	0.263**	
	(1.024)	(0.496)	(0.940)	(1.372)	(0.115)	
Effect Size (SD)	*0.59*	*0.5*	*0.33*	*0.32*	*0.26*	
Lango	2.246***	-0.464***	0.492*	1.342**	0.101***	
	(0.724)	(0.148)	(0.276)	(0.532)	(0.029)	
Effect Size (SD)	*0.27*	*0.29*	*0.15*	*0.20*	*0.22*	
Ateso	1.661	0.028	0.849**	0.980***	0.107***	
	(1.380)	(0.453)	(0.374)	(0.314)	(0.033)	
Effect Size (SD)	--	--	*0.19*	*0.22*	*0.23*	
English	2.345***	-0.144	1.197***	1.654**	0.086**	0.143
	(0.606)	(0.157)	(0.396)	(0.682)	(0.036)	(0.209)
Effect Size (SD)	*0.26*	--	*0.19*	*0.17*	*0.15*	--

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All regressions include district and school fixed effects, individual controls and a constant term. Individual controls include age and sex of learner, dummy for age missing, score on household assets index, dummies for student living with both parents, someone at home reading to the student and language spoken at home. Regressions for English language include control for local language.

Overall, the impact of the school-level SHRP intervention is positive, especially in Luganda and Lango. The Luganda subgroup is the one that shows the strongest improvement due to the SHRP intervention, as SHRP had a positive impact on all literacy sub-tasks. The second language sub-group that showed significant improvements due to the intervention is the Lango sub-group, although we do observe a negative and statistically significant impact on word segmenting, a seemingly strange and unexplained finding that does not fit with the general trend related to other literacy subtasks in Lango which all showed that the treatment group outperformed the control group (although the effect for non-word decoding is quite small and borderline significant).

Impact of school-level intervention:

- Positive impact on Luganda, Lango and English letter sound knowledge.
- No impact on R/R and Ateso letter sound knowledge.
- Positive impact on Luganda word segmenting, negative impact on Lango word segmenting. No impact on other languages word segmenting.
- Positive impact on non-word decoding, oral reading fluency and reading comprehension for all languages
- No impact on English receptive vocabulary

Among students in the R/R subgroup we find a positive impact in non-word decoding, oral reading fluency and reading comprehension but, perhaps surprisingly, no program impact on the more basic skills such as letter sound knowledge and phonemic awareness. Given that skills related to alphabetics (letter sound knowledge and phonemic awareness) are in theory foundational for decoding words[10], one would expect to first see improvements in these basic skills. Something similar occurs for Ateso language; there are no program effects on basic skills but we observe some positive effects on non-word decoding, oral fluency and reading comprehension although the effects are quite small.

Among Lango learners, there is a positive effect on letter sound scores, a negative effect on segmenting and almost zero effect on non-word decoding. Oral fluency shows a modest improvement due to the program.

Most of the improvements we see in English subtasks scores are due to the contribution of the Luganda learners, who showed relatively larger positive effects compared to other languages. In general the effects of the program among the Luganda students are the largest for scores in both languages, Luganda and English. When English scores are analyzed by language sub-group, we generally find no positive impact of SHRP on English scores of students from the other languages.

At this stage, it is unclear why impacts in Luganda are larger than the other local languages. Discussions with stakeholders, including USAID and the IP may help uncover possible reasons regarding this finding. It is possible that the intervention was implemented differently in Luganda speaking districts than in other areas (in terms of quality and effectiveness of teacher trainings and/or instructional materials for instance) or that the techniques taught were better suited to the Luganda language. We intend to delve deeper into these inquiries during upcoming stakeholder meetings and results dissemination workshops.

Although the analysis shows that SHRP has had a positive effect on a number of subtasks in several local languages, effect sizes are modest. Except for Luganda, letter sound knowledge and word segmenting, all other effect sizes are between 0.30 and 0.15 standard deviations, which is generally considered a small effect size. By comparison, other early grade reading interventions implemented by the same IP have yielded effect sizes of 0.80 standard deviations for oral reading fluency and reading comprehension after 2 years of intervention in Liberia[11] and effect sizes of 0.60 for Kiswahili letter-sound fluency, and 0.35 for Kiswahili oral reading fluency and reading comprehension after 2 years of intervention in Kenya[12]. Of course, we just provide this comparison as a reference. Different languages are not necessarily expected to show the same responses to the intervention.

Another finding is the lack of impact in English receptive vocabulary. The SHRP intervention focuses on mother-tongue based instruction in P1 through P3 with transition to English in P4, so it would not necessarily be surprising to see a lack of impact on higher level literacy skills in English for P2 students. However, the focus of SHRP in the early grades is precisely on building students' English oral language skills, especially receptive vocabulary. An explanation may be that the English receptive vocabulary subtask in the EGRA tool used by SHRP focuses on knowledge of body parts and spatial vocabulary which may not be the best gauge of receptive vocabulary, especially if teachers do not focus on these topics during their lessons.

It is also worth noting that while the intervention has had positive impacts on a number of subtasks, the absolute levels of competence remain very low. After two years of intervention, none of the subgroups reached an average of more than 13 letter sounds per minute; similarly none of the subgroups could read, on average, more than 7 words per minute in a connected text. While there are no benchmarks for a preferred fluency rate in these local languages, 7 words per minute would not in any case seem sufficient for being able to read with comprehension. Furthermore, we see that there are large numbers of learners that received the program who still cannot read a single word, and there are very few learners able to read more than 20 words from a paragraph. It is, therefore, not surprising that scores in reading comprehension are also extremely low. In the best performing subgroups, Luganda and R/R, the learners scored just over half a question on average, while in Lango and Ateso the treatment learners scored, on average, only 0.13 questions.

[10] Longitudinal studies have shown that phonemic awareness is highly predictive of decoding and that phonemic awareness is necessary, although not sufficient, for learning how to read. International Reading Association, 1998.

[11] Piper, B. & Korda, M. (2013). USAID/Liberia EGRA Plus: Final Program Evaluation. Retrieved from https://www.eddataglobal.org/countries/index.cfm?fuseaction=pubDetail&ID=541

[12] Piper, B. & Mugenda, A. (2014). USAID/Kenya Primary Math and Reading (PRIMR) Initiative – Endline Impact Evaluation Report. Retrieved from https://www.eddataglobal.org/countries/index.cfm?fuseaction=pubDetail&ID=661

ANNEX A. BALANCE AT BASELINE – TREATMENT VS. CONTROL

Table A.1. Treatment vs. Control, R/R Language

Variable	Treatment Mean (SE)	Control Mean (SE)	Difference (C - T)
Age of student	7.753	7.839	0.086
	(0.067)	(0.113)	
% of cases with missing age information	0.145	0.080	-0.065*
	(0.011)	(0.015)	
Gender of student (female = 1)	0.488	0.499	0.011
	(0.016)	(0.027)	
Number of assets	2.330	2.350	0.020
	(0.039)	(0.069)	
Lives with both parents (yes = 1)	0.733	0.724	-0.009
	(0.014)	(0.024)	
Does not live with mother (yes = 1)	0.139	0.134	-0.005
	(0.011)	(0.019)	
Someone reads to student at home (yes = 1)	0.545	0.538	-0.007
	(0.016)	(0.027)	
Attended preschool (yes = 1)	0.445	0.467	0.022
	(0.016)	(0.028)	
Student absent any day in the week prior to assessment (yes = 1)	0.545	0.512	-0.033
	(0.016)	(0.027)	
Teacher absent any day in the week prior to assessment (yes = 1)	0.420	0.405	-0.015
	(0.017)	(0.028)	
Gender of teacher (female = 1)	0.803	0.852	0.049
	(0.013)	(0.019)	
Teacher teaches in R/R (yes = 1)	0.782	0.763	-0.019
	(0.013)	(0.023)	
Speaks R/R at home (yes = 1)	0.857	0.896	0.039
	(0.011)	(0.017)	
R/R Letter sound score (max = 100)	3.070	3.389	0.319
	(0.171)	(0.271)	
R/R Word segmenting score (max = 10)	3.278	3.585	0.307
	(0.132)	(0.230)	
R/R Nonword decoding score (max = 50)	0.423	0.407	-0.016
	(0.070)	(0.123)	
R/R Oral reading fluency (max = 68)	0.344	0.303	-0.041
	(0.068)	(0.110)	
R/R Reading comprehension score (max = 5)	0.024	0.018	-0.006
	(0.006)	(0.009)	
R/R Listening comprehension	1.429	1.430	0.001
	(0.032)	(0.055)	
English Letter sound score (max = 100)	2.122	2.920	0.798*
	(0.132)	(0.249)	
English Word segmenting score (max = 10)	0.149	0.184	0.035
	(0.019)	(0.032)	
English Nonword decoding score (max = 50)	0.212	0.151	-0.061
	(0.045)	(0.070)	
English Oral reading fluency (max = 68)	0.281	0.191	-0.090
	(0.054)	(0.082)	
English Reading comprehension score (max = 5)	0.001	0.000	-0.001
	(0.001)	(0.000)	
English Receptive vocabulary score (max=20)	5.348	6.436	1.088**
	(0.092)	(0.146)	

*p<0.10, **p<0.05, ***p<0.001

Table A.2. Treatment vs Control, Luganda Language

Variable	Treatment Mean (SE)	Control Mean (SE)	Difference (C - T)
Age of student	6.899	7.251	0.352
	(0.058)	(0.138)	
% of cases with missing age information	0.136	0.123	-0.013
	(0.011)	(0.021)	
Gender of student (female = 1)	0.481	0.462	-0.019
	(0.015)	(0.033)	
Number of assets	3.029	2.941	-0.088
	(0.045)	(0.094)	
Lives with both parents (yes = 1)	0.463	0.419	-0.044
	(0.015)	(0.032)	
Does not live with mother (yes = 1)	0.299	0.339	0.040
	(0.014)	(0.031)	
Someone reads to student at home (yes = 1)	0.546	0.538	-0.008
	(0.015)	(0.033)	
Attended preschool (yes = 1)	0.744	0.689	-0.055
	(0.014)	(0.030)	
Student absent any day in the week prior assessment (yes = 1)	0.532	0.557	0.025
	(0.016)	(0.033)	
Teacher absent any day in the week prior to assessment (yes = 1)	0.482	0.538	0.056
	(0.016)	(0.033)	
Gender of teacher (female = 1)	0.966	0.911	-0.055
	(0.006)	(0.019)	
Teacher teaches in Luganda (yes = 1)	0.573	0.597	0.024
	(0.015)	(0.032)	
Speaks Luganda at home (yes = 1)	0.772	0.869	0.097***
	(0.013)	(0.022)	
Luganda Letter sound score (max = 100)	2.420	1.852	-0.568
	(0.157)	(0.259)	
Luganda Word segmenting score (max = 10)	2.351	3.129	0.778
	(0.115)	(0.272)	
Luganda Nonword decoding score (max = 50)	0.200	0.140	-0.060
	(0.044)	(0.078)	
Luganda Oral reading fluency (max = 68)	0.211	0.140	-0.071
	(0.042)	(0.079)	
Luganda Reading comprehension score (max = 5)	0.014	0.000	-0.014***
	(0.004)	(0.000)	
Luganda Listening comprehension	0.910	0.987	0.077
	(0.028)	(0.063)	
English Letter sound score (max = 100)	2.500	1.547	-0.953
	(0.167)	(0.227)	
English Word segmenting score (max = 10)	0.418	0.538	0.120
	(0.040)	(0.114)	
English Nonword decoding score (max = 50)	0.379	0.178	-0.201
	(0.062)	(0.100)	
English Oral reading fluency (max = 68)	0.840	0.483	-0.357
	(0.090)	(0.136)	
English Reading comprehension score (max = 5)	0.006	0.013	0.007
	(0.002)	(0.009)	
English Receptive vocabulary score (max=20)	7.538	6.775	-0.763
	(0.121)	(0.248)	

*p<0.10, **p<0.05, ***p<0.001

Table A.3. Treatment vs. Control, Langa Language

Variable	Treatment Mean (SE)	Control Mean (SE)	Difference (C - T)
Age of student	7.585	7.562	-0.023
	(0.044)	(0.079)	
% of cases with missing age information	0.091	0.106	0.015
	(0.008)	(0.016)	
Gender of student (female = 1)	0.498	0.503	0.005
	(0.014)	(0.025)	
Number of assets	2.511	2.497	-0.014
	(0.030)	(0.052)	
Lives with both parents (yes = 1)	0.665	0.711	0.046
	(0.013)	(0.023)	
Does not live with mother (yes = 1)	0.143	0.129	-0.014
	(0.010)	(0.017)	
Someone reads to student at home (yes = 1)	0.296	0.313	0.017
	(0.013)	(0.024)	
Attended preschool (yes = 1)	0.433	0.340	-0.093**
	(0.014)	(0.024)	
Student absent any day in the week prior assessment (yes = 1)	0.480	0.440	-0.040*
	(0.014)	(0.025)	
Teacher absent any day in the week prior to assessment (yes = 1)	0.349	0.334	-0.015
	(0.014)	(0.025)	
Gender of teacher (female = 1)	0.630	0.508	-0.122
	(0.014)	(0.026)	
Teacher teaches in Langa (yes = 1)	0.726	0.765	0.039
	(0.013)	(0.022)	
Speaks Langa at home (yes = 1)	0.907	0.902	-0.005
	(0.008)	(0.015)	
Langa Letter sound score (max = 100)	0.785	0.693	-0.092
	(0.086)	(0.160)	
Langa Word segmenting score (max = 10)	0.021	0.005	-0.016
	(0.008)	(0.004)	
Langa Nonword decoding score (max = 50)	0.016	0.016	0.000
	(0.012)	(0.016)	
Langa Oral reading fluency (max = 68)	0.064	0.031	-0.033
	(0.022)	(0.013)	
Langa Reading comprehension score (max = 5)	0.002	0.000	-0.002
	(0.002)	(0.000)	
Langa Listening comprehension	1.854	1.784	-0.070
	(0.029)	(0.053)	
English Letter sound score (max = 100)	0.653	0.430	-0.223
	(0.076)	(0.077)	
English Word segmenting score (max = 10)	0.653	0.430	-0.223
	(0.076)	(0.077)	
English Nonword decoding score (max = 50)	0.048	0.013	-0.035
	(0.012)	(0.013)	
English Oral reading fluency (max = 68)	0.010	0.008	-0.002
	(0.010)	(0.008)	
English Reading comprehension score (max = 5)	0.031	0.005	-0.026
	(0.016)	(0.004)	
English Receptive vocabulary score (max=20)	0.000	0.000	0.000
	(0.000)	(0.000)	

*p<0.10, **p<0.05, ***p<0.001

Table A.4. Treatment vs. Control, Ateso Language

Variable	Treatment Mean (SE)	Control Mean (SE)	Difference (C - T)
Age of student	7.097	6.616	
	(0.052)	(0.090)	-0.481***
% of cases with missing age information	0.245	0.270	
	(0.012)	(0.022)	0.025
Gender of student (female = 1)	0.496	0.500	
	(0.014)	(0.024)	0.004
Number of assets	2.229	2.484	
	(0.031)	(0.050)	0.255***
Lives with both parents (yes = 1)	0.665	0.722	
	(0.013)	(0.022)	0.057
Does not live with mother (yes = 1)	0.139	0.151	
	(0.010)	(0.018)	0.012
Someone reads to student at home (yes = 1)	0.347	0.347	
	(0.014)	(0.024)	0.000
Attended preschool (yes = 1)	0.266	0.287	
	(0.013)	(0.022)	0.021
Student absent any day in the week prior assessment (yes = 1)	0.540	0.543	
	(0.014)	(0.025)	0.003
Teacher absent any day in the week prior to assessment (yes = 1)	0.483	0.491	
	(0.014)	(0.026)	0.008
Gender of teacher (female = 1)	0.665	0.356	
	(0.014)	(0.023)	-0.309**
Teacher teaches in Ateso (yes = 1)	0.836	0.864	
	(0.010)	(0.017)	0.028
Speaks Ateso at home (yes = 1)	0.940	0.967	
	(0.007)	(0.009)	0.027
Ateso Letter sound score (max = 100)	1.810	1.782	
	(0.106)	(0.196)	-0.028
Ateso Word segmenting score (max = 10)	1.725	1.380	
	(0.094)	(0.147)	-0.345
Ateso Nonword decoding score (max = 50)	0.015	0.067	
	(0.008)	(0.042)	0.052
Ateso Oral reading fluency (max = 68)	0.049	0.050	
	(0.018)	(0.030)	0.001
Ateso Reading comprehension score (max = 5)	0.002	0.010	
	(0.001)	(0.006)	0.008
Ateso Listening comprehension	1.634	1.672	
	(0.026)	(0.045)	0.038
English Letter sound score (max = 100)	1.199	1.395	
	(0.089)	(0.153)	0.196
English Word segmenting score (max = 10)	0.285	0.156	
	(0.031)	(0.033)	-0.129
English Nonword decoding score (max = 50)	0.017	0.050	
	(0.010)	(0.031)	0.033
English Oral reading fluency (max = 68)	0.078	0.038	
	(0.023)	(0.022)	-0.040
English Reading comprehension score (max = 5)	0.002	0.000	
	(0.001)	(0.000)	-0.002
English Receptive vocabulary score (max=20)	2.369	1.801	
	(0.083)	(0.102)	-0.568*

*p<0.10, **p<0.05, ***p<0.001

ANNEX B. BALANCE AT BASELINE – TREATMENT VS. COMPARISON

Table 4. Treatment vs. Comparison, R/R Language

Variable	Treatment Mean (SE)	Comparison Mean (SE)	Difference (C - T)
Age of student	7.753	7.746	-0.007
	(0.067)	(0.116)	
% of cases with missing age information	0.145	0.073	-0.072**
	(0.011)	(0.015)	
Gender of student (female = 1)	0.488	0.484	-0.004
	(0.016)	(0.029)	
Number of assets	2.330	2.152	-0.178
	(0.039)	(0.076)	
Lives with both parents (yes = 1)	0.733	0.734	0.001
	(0.014)	(0.026)	
Does not live with mother (yes = 1)	0.139	0.145	0.006
	(0.011)	(0.021)	
Someone reads to student at home (yes = 1)	0.545	0.534	-0.011
	(0.016)	(0.030)	
Attended preschool (yes = 1)	0.445	0.486	0.041
	(0.016)	(0.030)	
Student absent any day in the week prior to assessment (yes = 1)	0.545	0.535	-0.010
	(0.016)	(0.030)	
Teacher absent any day in the week prior to assessment (yes = 1)	0.420	0.496	0.076*
	(0.017)	(0.031)	
Gender of teacher (female = 1)	0.803	0.702	-0.101
	(0.013)	(0.027)	
Teacher teaches in R/R (yes = 1)	0.782	0.723	-0.059
	(0.013)	(0.026)	
Speaks R/R at home (yes = 1)	0.857	0.830	-0.027
	(0.011)	(0.022)	
R/R Letter sound score (max = 100)	3.070	1.612	-1.458**
	(0.171)	(0.219)	
R/R Word segmenting score (max = 10)	3.278	3.609	0.331
	(0.132)	(0.254)	
R/R Nonword decoding score (max = 50)	0.423	0.000	-0.423**
	(0.070)	(0.000)	
R/R Oral reading fluency (max = 68)	0.344	0.000	-0.344**
	(0.068)	(0.000)	
R/R Reading comprehension score (max = 5)	0.024	0.000	-0.024**
	(0.006)	(0.000)	
R/R Listening comprehension	1.429	1.415	-0.014
	(0.032)	(0.056)	
English Letter sound score (max = 100)	2.122	1.311	-0.811**
	(0.132)	(0.185)	
English Word segmenting score (max = 10)	0.149	0.138	-0.011
	(0.019)	(0.030)	
English Nonword decoding score (max = 50)	0.212	0.010	-0.202**
	(0.045)	(0.010)	
English Oral reading fluency (max = 68)	0.281	0.007	-0.274***
	(0.054)	(0.005)	
English Reading comprehension score (max = 5)	0.001	0.000	-0.001
	(0.001)	(0.000)	
English Receptive vocabulary score (max=20)	5.348	5.367	0.019
	(0.092)	(0.157)	

*p<0.10, **p<0.05, ***p<0.001

Table 5. Treatment vs Comparison, Luganda Language

Variable	Treatment Mean (SE)	Comparison Mean (SE)	Difference (C - T)
Age of student	6.899	7.121	0.222
	(0.058)	(0.110)	
% of cases with missing age information	0.136	0.108	-0.028
	(0.011)	(0.018)	
Gender of student (female = 1)	0.481	0.487	0.006
	(0.015)	(0.029)	
Number of assets	3.029	2.710	-0.319**
	(0.045)	(0.078)	
Lives with both parents (yes = 1)	0.463	0.490	0.027
	(0.015)	(0.029)	
Does not live with mother (yes = 1)	0.299	0.284	-0.015
	(0.014)	(0.026)	
Someone reads to student at home (yes = 1)	0.546	0.532	-0.014
	(0.015)	(0.029)	
Attended preschool (yes = 1)	0.744	0.671	-0.073
	(0.014)	(0.027)	
Student absent any day in the week prior assessment (yes = 1)	0.532	0.513	-0.019
	(0.016)	(0.029)	
Teacher absent any day in the week prior to assessment (yes = 1)	0.482	0.505	0.023
	(0.016)	(0.029)	
Gender of teacher (female = 1)	0.966	0.850	-0.116
	(0.006)	(0.020)	
Teacher teaches in Luganda (yes = 1)	0.573	0.719	0.146**
	(0.015)	(0.026)	
Speaks Luganda at home (yes = 1)	0.772	0.735	-0.037
	(0.013)	(0.025)	
Luganda Letter sound score (max = 100)	2.420	1.454	-0.966**
	(0.157)	(0.204)	
Luganda Word segmenting score (max = 10)	2.351	2.467	0.116
	(0.115)	(0.227)	
Luganda Nonword decoding score (max = 50)	0.200	0.176	-0.024
	(0.044)	(0.076)	
Luganda Oral reading fluency (max = 68)	0.211	0.196	-0.015
	(0.042)	(0.077)	
Luganda Reading comprehension score (max = 5)	0.014	0.003	-0.011**
	(0.004)	(0.003)	
Luganda Listening comprehension	0.910	0.716	-0.194**
	(0.028)	(0.048)	
English Letter sound score (max = 100)	2.500	1.075	-1.425***
	(0.167)	(0.159)	
English Word segmenting score (max = 10)	0.418	0.294	-0.124
	(0.040)	(0.069)	
English Nonword decoding score (max = 50)	0.379	0.206	-0.173
	(0.062)	(0.098)	
English Oral reading fluency (max = 68)	0.840	0.471	-0.369
	(0.090)	(0.151)	
English Reading comprehension score (max = 5)	0.006	0.003	-0.003
	(0.002)	(0.003)	
English Receptive vocabulary score (max=20)	7.538	6.062	-1.476**
	(0.121)	(0.184)	

*p<0.10, **p<0.05, ***p<0.001

Table 6. Treatment vs. Comparison, Lango Language

Variable	Treatment Mean (SE)	Comparison Mean (SE)	Difference (C - T)
Age of student	7.585	7.687	0.102
	(0.044)	(0.092)	
% of cases with missing age information	0.091	0.132	0.041
	(0.008)	(0.018)	
Gender of student (female = 1)	0.498	0.497	-0.001
	(0.014)	(0.026)	
Number of assets	2.511	2.069	-0.442***
	(0.030)	(0.057)	
Lives with both parents (yes = 1)	0.665	0.626	-0.039
	(0.013)	(0.025)	
Does not live with mother (yes = 1)	0.143	0.148	0.005
	(0.010)	(0.019)	
Someone reads to student at home (yes = 1)	0.296	0.247	-0.049
	(0.013)	(0.023)	
Attended preschool (yes = 1)	0.433	0.318	-0.115**
	(0.014)	(0.025)	
Student absent any day in the week prior assessment (yes = 1)	0.480	0.488	0.008
	(0.014)	(0.026)	
Teacher absent any day in the week prior to assessment (yes = 1)	0.349	0.401	0.052
	(0.014)	(0.027)	
Gender of teacher (female = 1)	0.630	0.219	-0.411**
	(0.014)	(0.025)	
Teacher teaches in Lango (yes = 1)	0.726	0.728	0.002
	(0.013)	(0.023)	
Speaks Lango at home (yes = 1)	0.907	0.879	-0.028
	(0.008)	(0.017)	
Lango Letter sound score (max = 100)	0.785	0.885	0.100
	(0.086)	(0.160)	
Lango Word segmenting score (max = 10)	0.021	0.003	-0.018
	(0.008)	(0.003)	
Lango Nonword decoding score (max = 50)	0.016	0.027	0.011
	(0.012)	(0.020)	
Lango Oral reading fluency (max = 68)	0.064	0.121	0.057
	(0.022)	(0.035)	
Lango Reading comprehension score (max = 5)	0.002	0.000	-0.002
	(0.002)	(0.000)	
Lango Listening comprehension	1.854	1.901	0.047
	(0.029)	(0.053)	
English Letter sound score (max = 100)	0.653	0.470	-0.183
	(0.076)	(0.094)	
English Word segmenting score (max = 10)	0.653	0.470	-0.183
	(0.076)	(0.094)	
English Nonword decoding score (max = 50)	0.048	0.011	-0.037*
	(0.012)	(0.007)	
English Oral reading fluency (max = 68)	0.010	0.036	0.026
	(0.010)	(0.030)	
English Reading comprehension score (max = 5)	0.031	0.054	0.023
	(0.016)	(0.019)	
English Receptive vocabulary score (max=20)	0.000	0.000	0.000
	(0.000)	(0.000)	

*p<0.10, **p<0.05, ***p<0.001

Table 7. Treatment vs. Comparison, Ateso Language

Variable	Treatment Mean (SE)	Comparison Mean (SE)	Difference (C - T)
Age of student	7.097	6.773	-0.324**
	(0.052)	(0.065)	
% of cases with missing age information	0.245	0.261	0.016
	(0.012)	(0.022)	
Gender of student (female = 1)	0.496	0.497	0.001
	(0.014)	(0.025)	
Number of assets	2.229	2.394	0.165*
	(0.031)	(0.061)	
Lives with both parents (yes = 1)	0.665	0.761	0.096***
	(0.013)	(0.021)	
Does not live with mother (yes = 1)	0.139	0.102	-0.037*
	(0.010)	(0.015)	
Someone reads to student at home (yes = 1)	0.347	0.401	0.054
	(0.014)	(0.025)	
Attended preschool (yes = 1)	0.266	0.187	-0.079**
	(0.013)	(0.020)	
Student absent any day in the week prior assessment (yes = 1)	0.540	0.589	0.049
	(0.014)	(0.025)	
Teacher absent any day in the week prior to assessment (yes = 1)	0.483	0.487	0.004
	(0.014)	(0.026)	
Gender of teacher (female = 1)	0.665	0.772	0.107
	(0.014)	(0.021)	
Teacher teaches in Ateso (yes = 1)	0.836	0.891	0.055
	(0.010)	(0.016)	
Speaks Ateso at home (yes = 1)	0.940	0.980	0.040**
	(0.007)	(0.007)	
Ateso Letter sound score (max = 100)	1.810	1.365	-0.445*
	(0.106)	(0.186)	
Ateso Word segmenting score (max = 10)	1.725	1.551	-0.174
	(0.094)	(0.171)	
Ateso Nonword decoding score (max = 50)	0.015	0.000	-0.015*
	(0.008)	(0.000)	
Ateso Oral reading fluency (max = 68)	0.049	0.000	-0.049**
	(0.018)	(0.000)	
Ateso Reading comprehension score (max = 5)	0.002	0.000	-0.002
	(0.001)	(0.000)	
Ateso Listening comprehension	1.634	1.642	0.008
	(0.026)	(0.046)	
English Letter sound score (max = 100)	1.199	0.970	-0.229
	(0.089)	(0.124)	
English Word segmenting score (max = 10)	0.285	0.673	0.388
	(0.031)	(0.114)	
English Nonword decoding score (max = 50)	0.017	0.000	-0.017*
	(0.010)	(0.000)	
English Oral reading fluency (max = 68)	0.078	0.041	-0.037
	(0.023)	(0.029)	
English Reading comprehension score (max = 5)	0.002	0.000	-0.002
	(0.001)	(0.000)	
English Receptive vocabulary score (max=20)	2.369	2.147	-0.222
	(0.083)	(0.124)	

*p<0.10, **p<0.05, ***p<0.001